International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 5 Issue 4, May-June 2021 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Automatic Covid-19 Infected Chest X-Ray Image Classification using Support Vector Machine

Md. Abdul Matin[1], Abdur Rahman[2], S M Abdullah Al Shuaeb[3], Anwar Hossen[4]

[1]Lecturer (Physics), [2]Workshop Super (Mechanical), [3]Instructor (Tech), [4]Assistant Controller (Examination),
[1]Begum Rokeaya Girls School and College Gurudaspur, Natore, Bangladesh
[2,3]Tangail Polytechnic Institute, Tangail, Bangladesh
[4]Rabindra University, Shahjadpur, Bangladesh

ABSTRACT

The recent coronavirus disease (COVID-19) is extending very speedily over the world for the sake of its very infectious nature and is announced nationwide by the world health organization (WHO). The (COVID-19) is a group of coronavirus that has caused panic all over the world. It enters people through the sneezing and coughing of the infected person and weakens the person and it then slowly infects the affected person's lungs. In this study, we have classified the chest X-Ray images like Covid-19 infected chest images or normal chest images. Classifying the chest X-Ray images is hard and time-consuming work for human beings. Hence, an automatic Covid-19 infected chest X-Ray image or normal chest classification tool is very useful even for experience humans to classify a lot of chest X-Ray images. For that, we have proposed a new machine learning technique to automatically classify the chest Covid-19 infected X-Ray images or normal chest images. Hence, we have used a Machine learning (ML) model like Support Vector Machine (SVM) to classify (Covid-19) infected chest images and normal chest images. For this work, at first, we have preprocessed the chest X-Ray image. Then we have extracted the distinct features from the chest X-Ray images. After that, these features have trained into Machine Learning (ML) algorithm and finally classify these images into the category. From the experiment, The Support Vector Machine (SVM) models achieving an accuracy of up to 93.1%.

KEYWORDS: *Machine Learning (ML), Support Vector Machine (SVM), Confusion Matrix (CM), Covid-19, Chest X-Ray, Image Processing (IP)*

How to cite this paper: Md. Abdul Matin | Abdur Rahman | S M Abdullah Al Shuaeb | Anwar Hossen "Automatic Covid-19 Infected Chest X-Ray Image Classification using Support Vector Machine" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-5 | Issue-4, June 2021, pp.413-418, URL: www.ijtsrd.com/papers/ijtsrd41283.pdf

IJTSRD41283

1. INTRODUCTION

Covid-19 has become a burning issue in today's world. This virus has created a difficult situation all over the world. The outbreak began in late December 2019 in Wan, Hubei Province, China, and has since spread to 218 regions around the world, including Bangladesh. On March 8, 2020, the first case of the virus was identified in the country and ten days later, on March 18, the first person died of the virus, then the rate of infection gradually increased in Bangladesh[1].In the last two to three months of last year, the corona infestation and mortality graph in Bangladesh was very low. But in February and the end of March this year, there was a sudden increase in corona infestation and mortality at a time when people were not as alert. Due to which the number of patients infected with corona is gradually increasing in the big cities of Bangladesh. According to the Department of Health, as of May 11, 2021, the number of patients suffering from compassion in Bangladesh is over 7 lakh 76 thousand and the number of deaths due to compassion is 12005. The human body suffers from various types of damage when it is infected with the Covid-19, most notably the human airways and lungs.

Health experts have already said that these people have been infected with the Covid-19, their lungs have been largely destroyed and they will never fully recover. Covid-19 lungs are severely damaged and the most common symptoms are severe shortness of breath, cough, and fatigue. An X-ray of an infected person's chest reveals the condition of the person's lungs and how much damage has been done. Over the last year, there has been much research on Covid-19 chest X-Ray images in the world among the research includes X-Ray image classification, lung image, tumor classification, blood cell detection, etc. There is currently a lot of research being done on coronavirus infected lung images using machine learning algorithms. In recent years, image processing plays an important role in the part of machine learning[2]. Image processing (IP) means fetching necessary knowledge from the image. The X-Ray image classification task much like general image classification like a cat, dog, cow, etc[3]. Presently, X-Ray chest image classification is a significant thing to identify lung infection type or situation of the lung and detects their levels.

In this work, a novel technique is illustrated which is the organization of one classification model, Support Vector Machine (SVM) with various distinct sets of features[4]. The features learning are average red, green, blue color, hue, saturation, values color, and horizontal and vertical contrast, horizontal and vertical correlation, horizontal and vertical energy, horizontal and vertical homogeneity,

gray-level co-occurrence matrix (GLCM), and compost automatic various features for SVM algorithm and their convention. These features have complied to learn the machine learning (ML) models for classifying the X-Ray chest image.

In our research work, our image dataset contained 140 X-Ray images that are categorized into two classes of chests namely normal chest and covid-19 infected chest images. From the experimental result, we have investigated that the SVM algorithm classifies the normal chest image and Covid-19 chest image. From our observation, the Support Vector Machine (SVM) model illustrates the accuracy is (93.9%).

The rest of this paper is as follows. Section 2 describes the literature review. Section 3 represents the dataset and methodology. Section 4 depicts the results Section 5 displays the conclusion.

2. Literature Review

Saurabh Kumar et al. in [5] tried to classify chest X-Ray images using deep learning. They have scaled all the images of the data set to a uniform size of 512×512. Here they have used 401 images for these classification problems. This dataset contains 401 images out of which 262 images are covid-19 negative and the rest of these images are covid-19 positive and they have achieved a high accuracy result. For a good result, the deep learning model required huge image data but they have used only 401 images sample for this classification problem. So it was the limitation of their research work.

The authors in [6] described medical imaging such as X-ray and Computed Tomography (CT) associated with the potential of Artificial Intelligence (AI) plays a vital role in siding the medical staff in the diagnosis process. These types of image classification they have used five deep learning algorithm namely (ResNet18, ResNet34, InceptionV3, InceptionResNetV2, and DenseNet161). In this study, they used two public datasets and the first dataset was the COVID-19 image data collection, consisting of 236 images of COVID-19, 12 images of COVID-19 and ARDS, 4 images of ARDS, 1 image of Chlamydophila, 1 image of Klebsiella, 2 images of Legionella, 12 images of Pneumocystis, 16 images of SARS, 13 images of Streptococcus and 5 images without any pathological findings and the second dataset was covid-19 X-Ray images. Here, they have used a deep learning model for classifying the covid-19 positive or negative case. The performance of the Deep Learning Algorithm depends on its data set which means the amount of data needed for good results but here they have used very few images for their work so the algorithm may provide bias results.

Shelke et al. in [7]were described the classification model to classify the X-Ray image. Here, they have used four classes of X-Ray images namely normal, pneumonia, tuberculosis (TB), and COVID-19. They have used a deep learning model used for the classification of pneumonia, TB, and normal is VGG16 with an accuracy of 95.9 %. But they don't explain how many images they've used for their research that is the drawback of their work.

Another CXR image classification task was described here[8]. Here, the authors have used the deep Residual Network (RESNET-50). In this classification, they have used four types of CXR image cases like healthy individuals, bacterial and viral pneumonia, and COVID-19 positives patients. The model performance metrics showed an accuracy of 99%. But a large amount of data are required for the deep learning algorithm for high accuracy but a small number of images were used for these purposes which is the limitation of their research.

The authors in [9] were described the rapid development in the area of Machine Learning (ML) and Deep Learning. Here, they have proposed intelligent systems to classify between Pneumonia and Normal patients. The proposes of the machine learning-based classification of the extracted deep feature using ResNet152 with COVID-19 and Pneumonia patients on chest X-ray images. SMOTE was used for balancing the imbalanced data points of COVID-19 and Normal patients. The model has achieved accuracy up to .97% on Random Forest and 97.4%n using XGBoost predictive classifiers.

3. Methodology and Dataset

In this part, the algorithm and datasets are described. The algorithm is used to classify covid-19 infected and normal chest X-Ray images. The algorithm includes SVM (Support Vector Machine). The chest X-Ray images of two levels of normal and covid-19 are provided as the dataset in the algorithm.

3.1. Proposed System

The proposed system block diagram is shown here. The block diagram is shown in Fig.3.1. The training images are resized, and many pre-processing are executed. So that the original images may carry many noise and errors. Then the contrast of the images is increased. Many features are extracted and then the machine learning model is applied to classify the images. When the training is finished, the accuracy of the model is calculated using the test images and confusion matrix.

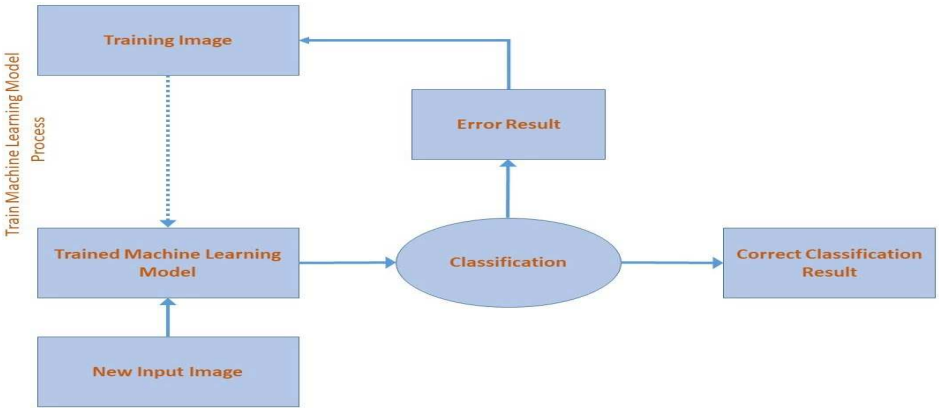

Fig. 3.1: Proposed System Block Diagram

The algorithm of the proposed system is illustrated in Algorithm 1. This algorithm represents the steps of the proposed model.

Algorithm 1: Covid-19 chest X-Ray image classification

1 Read training images from the dataset
2 Pre-process images
3 Train the model SVM to classify
4 Read test image
5 Apply the trained model to classify the test image
6 Control the classification_error
7 If (False_result> threshold)
Jump to step 1.
Else
achieve accuracy.

3.2. Dataset Description

In this work, we have collected many chest X-Ray images from different Medical colleges and private Hospitalsof the Mymensingh and Tangail districts in Bangladesh.

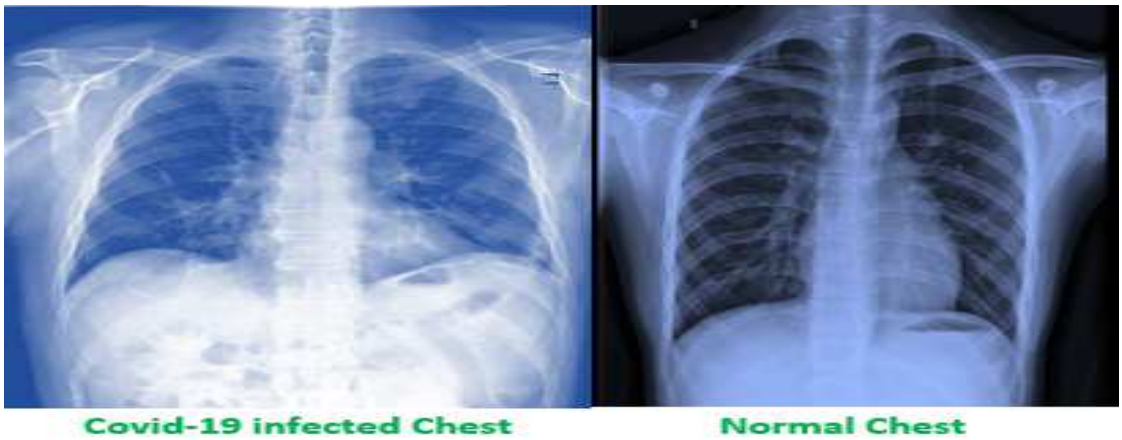

Fig 3.2: Two levels Chest X-Ray Images

These chest X-Ray images having two levels were taken from different Medical colleges and private Hospitals of the Mymensingh and Tangail, Dhaka, Natore districts in Bangladesh. The chest X-Ray images are then categorized into two classes based on two levels of covid-19 infected and normal chest depicted in the figure. 3.2. Each class has more than seventy images. Details of each class are shown in Table 3.1. These images are divided randomly into training (70%) and testing (30%) images.

Table 3.1: Details of Image Dataset

Class Name	Label	No. of image
Covid-19 Infected image	0	75
Normal images	1	79
Total Images		154

3.3. Image Pre-processing

Training images are collected from different places may be of various resolution and with noise. Therefore, image pre-processing is needed for reshaping the images and avoiding the noise. We have used Matlab*resize ()* function to resize the pictures to 350*180 pixels, *filter ()* function for smoothing, sharpening. Here, we have also used the *gray2rgb ()* function for eliminating the hue and saturation information, *gray_comatrix () function for using texture analysis of the images.*

Tab 3.2: Image Preprocessing Function

ML Model	Description	Use the Matlab Functions
SVM	Resize to 350 x 180 pixels	*resize ()*
	Filtering for smoothing, sharpening	*filter ()*
	The gray2rgb function for eliminating the hue and saturation information of the images.	*rgb2gray ()*
	Graycomatrix function for texture analysis of the images.	*graycomatrix ()*

3.4. Feature Extraction

Feature learning is the next processing step in image analysis. It can be used for images. Feature extraction is the measurable character of an image or object.

Tab 3.3: Feature Extraction Function of machine learning model

ML Model	Feature Name	Description	Matlab Functions
SVM	Average red color	The average value of all red pixels in the chest X-Ray image surface	*Rave=uint8 (mean ())*
	Average green color	The average value of all green pixels in the chest X-Ray image surface	*Gave=uint8 (mean ())*
	Average blue color	The average value of all blue pixels in the chest X-Ray image surface	*Bave=uint8 (mean ())*
	Average hue color	The average value of all hue pixels in HSV the chest X-Ray image surface	*Have=uint8 (mean ())*
	Average saturation color	The average value of all saturation pixels in HSV the chest X-Ray image surface	*Save=uint8 (mean ())*
	Average values color	The average value of all values pixels in HSV chest X-Ray image surface	*Save=uint8 (mean ())*
	Horizontal and Vertical Contrast	Find the local variation of the gray-level co-occurrence matrix	*Contrast ()*
	Horizontal and Vertical Correlation	Find the joint probability occurrence of the specified pixel pairs.	*Correlation ()*

3.5. Chest X-Ray Image Classification or prediction

It classifies the target chest X-Ray class into a predefined input image using a machine learning algorithm. After the finishing of the feature extraction process, the images are first trained through the machine learning SVM (Support Vector Machine) model, and when the finishes the training process, we have examined with a single image that is not training by the trained classifier. The testing classifier compares the new testing image with previously trained different image levels. When the testing image is equalized the levels train images the machine learning model classifies the target chest X-Ray image levels.

3.6. Machine Learning Models

Machine learning (ML) is the data analytical design that instructs computers to do what comes naturally to human study from knowledge. Machine learning (ML) models use the computational process to learn facts immediately from data without depending on a predetermined equation as an algorithm[10].The models adaptively progress their achievement as the number of samples gain able for studying increases. In this work, we have to study Support Vector Machine (SVM) machine learning algorithms to complete the automatic covid-19 infected chest image classification work.

3.7. Support Vector Machine (SVM) Model

Support vector machine is a supervised Machine Learning (ML) algorithm. It can be used for both classification and regression problems[11]. But it is popularly used for classification. The primary goal of the Support Vector Machine (SVM) is to find the hyper plane which divides the two classes of data.

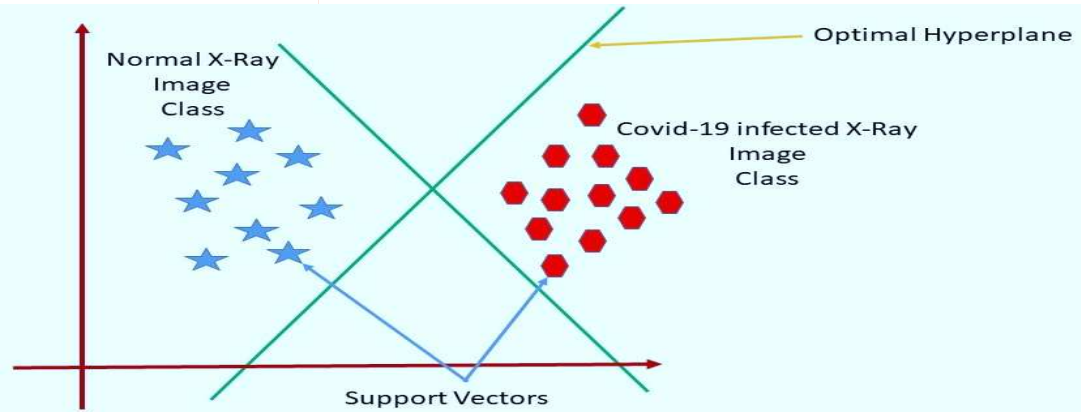

Fig.3.3: Internal Structure of Support Vector Machine (SVM)

In most cases, a support vector machine is used to classify the data. A hyper plane which partitions two class with maximum distance is called an optimal hyper plane. SVM is very skillful for the supervised classifier. In this work, the Support Vector Machine (SVM) algorithm is used for two types of chest X-Ray image classification shown in (Fig.3.3). An SVM model is a representation of the examples as points in the coordinate system, mapped so that the sample of the two classes are divided by a clear gap that is as wide as possible. Given a training set of two classes, G = { (xi, Yi), i = 1 … N} with a hyperplane $W^T\phi(x_i) + b = 0$, xi ϵ Rn and y ϵ {1, −1}, the support vector machine satisfies the following conditions:

$(W^T\phi(x_i) + b) \geq 1, if\ yi = 1,$ (3.5)

$(W^T\phi(x_i) + b) \leq -1, if\ yi = -1,$ (3.6)

Or equivalently,

$y_i(W^T\phi(x_i) + b) \geq 1, if\ i = 1,2,3,\ldots\ldots\ldots N$ (3.7)

Where ϕ is the function that maps training vector xi to the higher dimensional space when the data points are linearly separable. The distance from a point xi to the hyper plane is:

$$\frac{I(W^T\phi(x_i)+b)I}{||w||^2} \quad (3.8)$$

From the definition of SVM, the margin is $\frac{2}{||W||}$. Hence, the equation of hyper plane is $\frac{1}{2}||W||^2$ (3.9)

According to the saddle point of the Lagrange function, the solution of the above equation is,

$$L_{p1} = \frac{1}{2}||w||^2 - \sum_{i=1}^{n}\alpha_i[y_i(w^T\phi(x_i)+b)-1] \quad (3.10)$$

where α_i are the nonnegative Lang range multipliers. When the data is not separable, a new slack variable ξ_i is introduced and the optimization equation is

$$y_i(W^T\phi(x_i)+b) \geq 1-\xi_i \quad (3.11)$$

And the hyper plane equation is- $\frac{1}{2}||w||^2 + c\sum_{i=1}^{n}\xi_I$ (3.12)

where C is a positive constant parameter used as a penalty parameter for the error term. If the optimization of the support vector machine uses linear and radial basis function, then the equation is: $K(x_i, x_i) = x\frac{T}{i}x_j$

$$K(x_i, x_i) = expexp\left(-\gamma||x_i - x_i||^2\right), \gamma > 0 \quad (3.13)$$

Where γ is the kernel parameter. Figure. 3.3 represents the visualization of SVM. The main advantage of the support vector machine is that it is effective in high dimensional spaces and it also works well with a clear margin of separation. The primary drawback of the support vector machine is that it does not well perform when the data set is large. The support vector machine also low performance, the data set, is rowdy.

4. Result Analysis

In this part, we have discussed the classification performance of the support vector machine for these two classes chest X-Ray image prediction task. The Confusion Matrix (CM) is a table that mostly behaves to consider the achievement of a classification model on a set of testing data for which the true values are acquainted[12]. The Confusion Matrix (CM) itself is comparatively easy to realize for that many researchers use it.

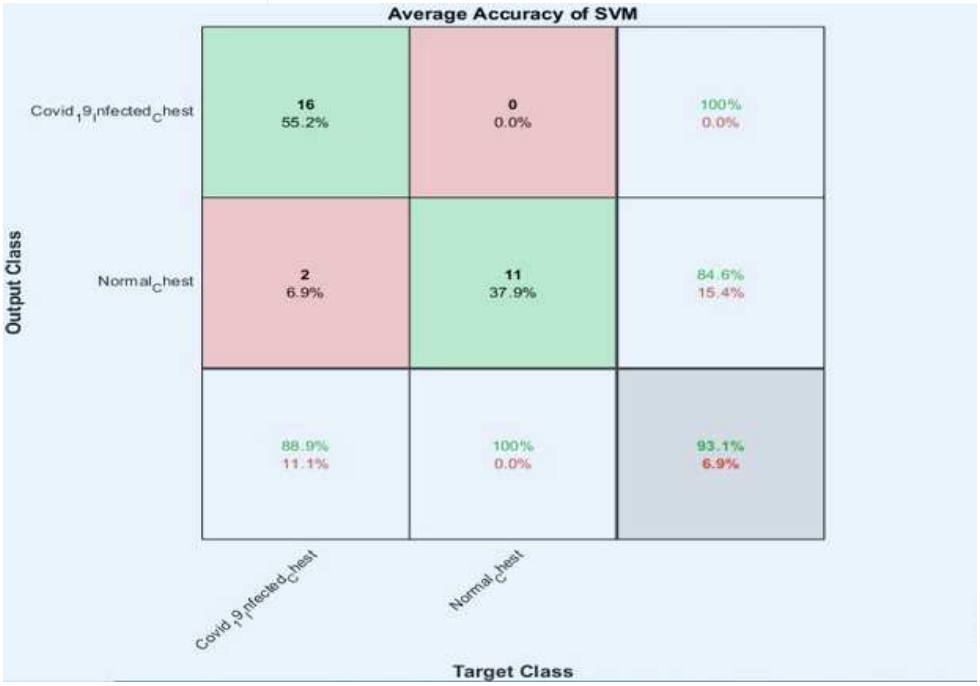

Fig. 4.1: Confusion Matrix of Support Vector Machine

In (Fig.4.1) we see that diagonally shaded boxes display the percent accuracy result of the SVM model. On the other side shaded box illustrated the percent of mistakes for the classification problem. The average accuracy of the support vector machine classifier is depicted for the classification problem is 93.1% that means the average achievement accuracy for the overall classifier with the best value is (93.1%). In common mistake for the overall classifier with the value is (6.9%). In this study, we see that's the covid-19_infected_chesthas displayed the highest and theNormal_chest illustrated the lowest classification accuracy respectively.

5. Conclusion

In the study, we have proposed a novel technic to classify Chest X-Ray images with two-level using the Machine Learning (ML) model. The raised system is used machine learning models to automatically classify the Covid-19 infected chest X-Ray image and Normal chest X-Ray image. Our proposed system includes three phases: Image pre-processing features learning or extraction, and classification. Image Pre-processing means resizing the image, noise avoiding. Then, we have extracted distinct features like RGB color, HSV color, horizontal and vertical contrast, etc are extracted. In fine, the classification part can be performed after the feature vectors are propagated for every image. The Support vector machine (SVM) classifier is used for the classification problem. Our proposed system has been performed evaluated using

154Chest X-Ray images. From among these images, we have used 70 % images for training and 30% images for testing randomly. From the experimental result, we have achieved accuracy up to 93.1%.

ACKNOWLEDGMENT

The authors wish to thank BegumRokeayaGirlsSchool and college Gurudaspur, Natore, Bangladesh. The authors also thank its research lab for their helpful guidance and useful supports.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

References

[1] S M Abdullah Al Shuaeb | Md. Kamruzaman | Mohammad Al-Amin, "COVID 19 Outbreak Prediction and Forecasting in Bangladesh using Machine Learning Algorithm," *Int. J. Trend Sci. Res. Dev.*, vol. 5, no. 1, pp. 829–835, 2020.

[2] A. V Galphade and K. H. Walse, "Supervised Learning Approach for Flower Images using Color , Shape and Texture Features," *Int. Res. J. Eng. Technol.*, vol. 6, no. 5, pp. 5682–5688, 2019.

[3] C. Chen, Q. Yan, M. Li, and J. Tong, "Classification of blurred flowers using convolutional neural networks," *ACM Int. Conf. Proceeding Ser.*, pp. 71–74, 2019, doi: 10.1145/3342999.3343006.

[4] I. Patel and S. Patel, "Flower identification and classification using computer vision and machine learning techniques," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 277–285, 2019, doi: 10.35940/ijeat.E7555.088619.

[5] S. Kumar, S. Mishra, and S. K. Singh, "Deep transfer learning-based COVID-19 prediction using chest X-rays," *medRxiv*, no. September 2003, 2020, doi: 10.1101/2020.05.12.20099937.

[6] S. Chatterjee *et al.*, "Exploration of Interpretability Techniques for Deep Covid-19 Classification Using Chest X-Ray Images," *arXiv*, 2020.

[7] A. Shelke *et al.*, "Chest X-ray classification using Deep learning for automated COVID-19 screening," *medRxiv*, no. December 2019, 2020, doi: 10.1101/2020.06.21.20136598.

[8] Z. Tang *et al.*, "Severity assessment of coronavirus disease 2019 (COVID-19) Using quantitative features from chest CT images," *arXiv*, vol. 2019, pp. 1–18, 2020.

[9] R. Kumar *et al.*, "Accurate Prediction of COVID-19 using Chest X-Ray Images through Deep Feature Learning model with SMOTE and Machine Learning Classifiers," *medRxiv*, pp. 1–10, 2020, doi: 10.1101/2020.04.13.20063461.

[10] Y. Baştanlar and M. Özuysal, "Introduction to machine learning," *Methods Mol. Biol.*, vol. 1107, pp. 105–128, 2014, doi: 10.1007/978-1-62703-748-8_7.

[11] H. Bhavsar and M. H. Panchal, "A Review on Support Vector Machine for Data Classification," *Int. J. Adv. Res. Comput. Eng. Technol.*, vol. 1, no. 10, pp. 2278–1323, 2012.

[12] D. Houcque, "Introduction To Matlab for Engineering Students," no. August, 2005.

Authors Profiles

SN	Name and Designation	Qualification and Experiences	Photo
01	Md. Abdul Matin Lecturer of Physics Begum Rokeaya Girls School and College Gurudaspur, Natore, Bangladesh Under the Ministry of Education, Bangladesh. Email: fakirmatinphysics@gmail.com	I am a Lecturer of Physics Department at Begum Rokeaya Girls School and College Gurudaspur, Natore, Bangladesh Under the Ministry of Education, Bangladesh. I have been in a great profession like teaching for about 9 years. I got my bachelor's degree from Rajshahi University (RU)Bangladesh. Even though I studied physics, but, I have a lot of interest in applied Physics and Electronics, computer engineering, especially machine learning, image processing, etc. I want to take my research program a long way	
02	Engr. Abdur Rahman Workshop Superintendent (Mechanical) Tangail Polytechnic Institute, Tangail. Directorate of Technical Education, Under Technical and Madrasha Education. Division of Ministry of Education, Bangladesh. Email: rahman9332@gmail.com	I am a teacher of the Mechanical Department at Tangail Polytechnic Institute, Tangail, Bangladesh. I have been in a great profession like teaching for about 9 years. I got my bachelor's degree from Dhaka University of Engineering and Technology (DUET), Bangladesh. Although I am a mechanical engineer, I have a lot of interest in computer engineering, especially machine learning, image processing, etc. I want to take my research program a long way.	
03	**S M Abdullah Al Shuaeb** Instructor(Tech), Computer Technology, Tangail Polytechnic Institute, Tangail. Directorate of Technical Education, Under Technical and Madrasha Education Division of Ministry of Education, Bangladesh. Email: nixon.cse28@gmail.com	I am an Instructor (Tech), Computer Technology, Tangail Polytechnic Institute, Tangail, Bangladesh. I have 8 years of experience in teaching. I received my B.Sc. Engineering degree in CSE from the University of Jatiya Kabi Kazi Nazrul Islam University(JKKNIU), Trishal, Mymensingh, and MS in CS from Bangladesh Agriculture University(BAU), Bangladesh respectively. My area of interest for research are Machine Learning, Computer Vision, Image Processing, Digital Signal Processing, and Bioinformatics. For the last three years, I am accelerating my research journey in the area of machine learning.	
04	**Anwar Hossen** Assistant Controller of Examinations Rabindra University, Bangladesh Shahjadpur, Sirajganj, Bangladesh Email: suhag.cse@gmail.com	I am Assistant Controller of Examinations Rabindra University, Shahjadpur, Sirajganj, Bangladesh. I have 5 years of experience in controller examination. I received my B.Sc. Engineering degree in CSE from the University of Jatiya Kabi Kazi Nazrul Islam University, Trishal, Mymensingh,. My area of interest for research are Machine Learning, Computer Vision, Image Processing, Digital Signal Processing, and Bioinformatics. For the last three years, I am accelerating my research journey in the area of machine learning.